Evidence of improving quality of reporting of randomized controlled trials in subfertility

Sofia Dias¹, Roseanne McNamee and Andy Vail

Biostatistics Group, University of Manchester, Manchester, UK

¹To whom correspondence should be addressed at: Biostatistics Group, University of Manchester, Stopford Building, Oxford Road, Manchester M13 9PT, UK. E-mail: sofia.dias@manchester.ac.uk

BACKGROUND: The quality of randomized controlled trials (RCTs) in subfertility and their suitability for inclusion in meta-analyses have been assessed in the past and found to be insufficient. Our aim was to assess whether this quality has improved over time, particularly since the publication of the Consolidated Standards of Reporting Trials (CONSORT) statement, and to assess what proportion of trials could be included in the meta-analyses of pregnancy outcomes such as those included in Cochrane Reviews. METHODS: A selection of subfertility trials published in 1990, 1996 and 2002 was collected from the Cochrane Menstrual Disorder and Subfertility Group (MDSG) database. Only trials published in English as full journal articles, claiming to be randomized and reporting on pregnancy outcomes, were included. RESULTS: One hundred and sixty-four trials met our inclusion criteria. Twenty-four (15%) were found not to be randomized, despite claims, and only 10 trials (6%) provided adequate details on the methods of randomization and allocation concealment. Of these, only three had sufficient details extractable to allow for an intention-to-treat analysis of the outcome ‘live birth’. CONCLUSIONS: Although an improvement in some subfertility-specific issues was observed, the quality of reporting of RCTs still needs to improve to make them suitable for inclusion in meta-analyses such as those in the Cochrane Library.

Key words: methodological quality/statistics/subfertility/systematic review

Introduction

The Consolidated Standards of Reporting Trials (CONSORT) statement (Begg et al., 1996) was published in response to perceived bad quality of reporting of randomized controlled trials (RCTs) (Schulz et al., 1995) and revised in 2001 (Altman et al., 2001; Moher et al., 2001b). Concealed randomization of patients to groups and analysis of the data based on the ‘intention-to-treat’ (ITT) principle are cornerstone recommendations. Recent analysis has demonstrated that despite improvements, poor reporting persisted in 2000 in PubMed indexed trials (Chan and Altman, 2005) and in 2001 in the leading subfertility journals (Vail and Gardener, 2003). The latter review was limited to the two main subspecialty journals, which may publish better quality papers, and focused on a single year, therefore providing no information on how quality might have changed over time.

Systematic reviews of the evidence from RCTs usually include a formal statistical summary, or ‘meta-analysis’ of the data retrieved from the original trial publications. For patients involved in fertility studies, the outcome of chief interest is birth itself, rather than an intermediate step in the fertilization process. For this reason, only trials that report pregnancy outcomes for all women were analysed. Even if a trial is otherwise of good quality, poor reporting and presentation of data—to which fertility trials may be especially prone—may mean that the trial has to be excluded.

In addition to general concerns in common with other areas of research, there are particular features of subfertility data that, if not accounted for properly in design and analysis, generate statistical errors that, in turn, may lead to inappropriate conclusions (Daya, 2003). Four areas are of specific concern:

(i) Repeated cycles have been used to define ‘treatment periods’ for crossover studies in which participants are initially allocated at random but then follow a defined sequence of interventions, often alternating between treatments, until the success or the end of study. There has been ongoing debate in the subfertility literature regarding the use of such designs (Daya, 1993; Cohlen et al., 2004; Norman and Daya, 2004), including contradicting claims of how valid analyses can be obtained (Eijckemans et al., 2002; McDonnell et al., 2004). However, there is agreement in the statistical literature that data from a crossover design cannot provide valid inferences using standard statistical tests, when a participant’s outcome in one period affects participation in subsequent periods (Senn, 1993). When a crossover design is used, only first-stage data (before crossover) are suitable for inclusion in meta-analyses with parallel group studies and these should be clearly presented.
(i) Randomization should be performed at the latest possible stage before the study intervention is administered, so as to minimize possible dropouts or exclusions. Often in subfertility trials, a sequence of therapies or interventions is given before the study intervention. Failure to respond satisfactorily at preliminary stages will mean that the study intervention cannot be administered but, if already randomized, the patient should still be included in the report of results. Ideally, for trials of subsequent interventions, only patients who have responded to the preliminary interventions should be randomized, e.g. if the intervention is the number of embryos to transfer, only consenting patients who reach the transfer stage (after a number of pre-randomization treatments not relevant to the intervention) should be randomized. In any case, the timing of randomization and reason for exclusions should be clearly reported.

(ii) There are multiple stages of subfertility outcome from failure of stimulation through to live birth, via successful fertilization and culture of high-quality embryos, implantation and then biological, clinical and ‘ongoing’ pregnancy. Ultimate success depends on progressing through each previous stage. Although information on these rates may be of some interest, standard methods of statistical analysis, based on the assumptions of independence between observations, do not apply to such data because multiple observations on the same woman are likely to be correlated.

(iii) Subfertility data are generally hierarchical in nature. For each woman, the data can have a tree-like structure: there may be multiple cycles, within each of which there are multiple embryos, arising from multiple oocytes. Typical definitions of pregnancy, fertilization and implantation rates, per cycle, per oocyte or per embryo, respectively, treat two observations in one woman as equivalent to one observation in each of two women. Although information on these rates may be of some interest, standard methods of statistical analysis, based on the assumption of independence between observations, do not apply to such data because multiple observations on the same woman are likely to be correlated.

(iv) Patient flow: the number of patients randomized, excluded and analysed should always be made clear in reports. Knowing the number of women initially randomized to treatment groups, as well as the number of women with positive outcomes, is essential to carry out an ITT approach.

(v) There are multiple stages of subfertility, and what proportion of trials can be included in meta-analyses of pregnancy outcomes such as those included in Cochrane Reviews. In some cases, the number of women randomized and reported on any form of pregnancy (even if not as the main outcome) may be considered evidence of selective reporting. Although it may be acceptable to present some results in this form, it should not replace the presentation of the ‘per randomized woman’ results and should not be used in the main analysis. For example, when women are randomized to receiving multiple cycles of a given treatment, and the number of pregnancies per cycle are considered in the analysis, a unit-of-analysis error has occurred. We assessed whether such errors were present for any pregnancy outcome or adverse event and whether the correct results could be extracted from the information provided in the report.

(vi) Adverse events: the number of miscarriages and ectopic and multiple pregnancies are important adverse events in subfertility and should be reported even if none occurred. Failure to disclose any adverse events may be considered evidence of selective reporting.

Materials and methods

A selection of subfertility trials was collected from the Cochrane Menstrual Disorder and Subfertility Group (MDSG) database. The database includes all mentions of studies and RCTs in subfertility, published in several journals and Conference Proceedings (US Cochrane Centre, 2004). On 1 December 2004, 3455 references were available. The earliest available records were published in 1966 (three references), and the database was complete up to 2003. For practical reasons, we chose to remove all trials not published in English and to discard those references not published as full journal articles (such as abstracts from conferences and letters), because the description of methodology is necessarily restricted in such publications. After removing all such references, 1903 records remained for analysis.

We decided to review all trials in the database that purported to be randomized and reported on any form of pregnancy (even if not as the main outcome). We restricted our analysis to trials published in the years 1990, 1996 and 2002 to assess any changes in quality over time, and in particular to capture any differences in quality before the publication of CONSORT (1990 and 1996) and after the CONSORT guidelines had been available for sufficiently long for their impact to be noticeable (2002). Two hundred and sixty-three references were available for these years, all of which were retrieved and checked for eligibility. Ninety-nine references did not satisfy our inclusion criteria and were excluded: 5 references (5%) did not correspond to clinical trials (e.g. literature reviews and case studies); 44 references (44%) did not claim to be randomized or were not randomized at a level allowing comparison of pregnancy outcomes (e.g. randomization of embryos to intervention, then best embryos transferred regardless of randomization); 32 (32%) did not mention any pregnancy outcomes; and 18 references (18%) were not considered to report fertility trials, i.e. not all patients in the trial or in a pre-identified subgroup had live birth as main aim, although this may have been reported for some of the patients (e.g. study of the impact of a procedure on future fertility when not all patients will attempt to become pregnant or study of interventions to improve sperm count in healthy volunteers).

One hundred and sixty-four references were available for review. The quality of these references was assessed according to the Cochrane Handbook (Higgins and Green, 2005) and the CONSORT requirements, focusing on the following aspects:

(i) Randomization method.

(ii) Allocation concealment: we have assumed that when no information is present, allocation was not concealed. A distinction has also been made between trials where allocation was clearly adequately concealed and those where there is some mention of concealment, but it is unclear whether this was achieved adequately.

(iii) Trial design: crossover trials are unsuitable for ‘terminal event’ outcomes such as pregnancy and birth that preclude entry to subsequent phases.

(iv) Patient flow: the number of patients randomized, excluded and analysed should always be made clear in reports. Knowing the number of women initially randomized to treatment groups, as well as the number of women with positive outcomes, is essential to carry out an ITT approach.

(v) Unit-of-analysis errors: these errors occur when the unit of randomization is not the unit used as a denominator in the analysis and presentation of results. Although it may be acceptable to present some results in this form, it should not replace the presentation of the ‘per randomized woman’ results and should not be used in the main analysis. For example, when women are randomized to receiving multiple cycles of a given treatment, and the number of pregnancies per cycle are considered in the analysis, a unit-of-analysis error has occurred. We assessed whether such errors were present for any pregnancy outcome or adverse event and whether the correct results could be extracted from the information provided in the report.

(vi) Adverse events: the number of miscarriages and ectopic and multiple pregnancies are important adverse events in subfertility and should be reported even if none occurred. Failure to disclose any adverse events may be considered evidence of selective reporting (Hahn et al., 2000, 2002).

Results

The Cochrane MDSG database contained references to trials that have been published in 152 different journals. It was
decided that the two main fertility journals [Fertility & Sterility (F&S) and Human Reproduction (HR)] should be analysed separately, as they publish the bulk of the trials in this field, are perceived as high-quality journals and have been the subject of a previous quality review (Vail and Gardener, 2003). In the 3 years considered, F&S published 61 trials, HR 50 trials and the ‘other’ journals 53 trials.

The specialty of the main intervention was recorded as ‘Laboratory’ for trials where the intervention does not take place directly on the patient (e.g. different culture media for oocyte fertilization), ‘Medical’ for drug trials and ‘Surgical’ for trials where the intervention involves anaesthetics or a surgical procedure (e.g. different types of catheter for embryo transfer) (Table I). ‘Main intervention’ was defined as the intervention of chief interest in the study or the ‘new’ intervention, i.e. the reason the trial was conducted.

**Randomization method**

In this review, 81 trials (49%) were found not to have provided details on the randomization method. A further 24 (15%) had in fact allocated treatments in a non-random way, despite the ‘randomized’ claim. Of these, 10 used alternate treatment assignment and 11 used odd/even patient number, date of treatment, date of birth etc. to allocate patients to treatments. Three trials used totally *ad hoc* methods of treatment assignment (two studies allowed patients to choose their own preferred treatment and one chose controls as being the patients to whom treatment could not be applied due to weekends). More surgical trials used non-random methods of allocation as opposed to medical or laboratory trials. The proportion of trials using clearly non-random methods of allocation was similar for the 3 years considered (Figure 1). The proportion on trials describing an adequate randomization method increased from 16% in 1990 to 40% in 1996 and then 48% in 2002. HR published fewer non-randomized reports and had a higher proportion of trials accurately describing the allocation method than the other journals (Table II).

**Allocation concealment**

Allocation concealment is always possible and is an important indicator of the methodological quality of a randomized control trial. In the CONSORT checklist, not only are authors required to detail how the allocation sequence was generated, but also whether this sequence was concealed until the treatment was assigned. The most secure method of allocation concealment is third-party allocation, where someone not involved in the trial holds the allocation schedule. However, properly sealed, serially numbered, opaque envelopes containing the allocation schedule are also acceptable. Statements of the form ‘the doctor was blinded to treatment allocation’ were classified as ‘unclear concealment’.

Of all the trials assessed, only 17 (10%) had described adequate forms of allocation concealment and 11 of these were published in 2002 (Figure 1). Although the leading fertility journals publish most of the trials with adequate allocation concealment (of the 17 trials, F&S 8 and HR 7), the total number of clearly adequately concealed trials is disappointing: 17 of 164 (10%) (Table II). Even assuming that in reports classified as ‘unclear allocation’ was adequately concealed (albeit poorly reported), the number of concealed trials only rises to 39 of 164 (24%).

**Trial design (crossover trials)**

The number of published trials using crossover designs decreased considerably from 11 trials (24%) in 1990 to 8 (13%) in 1996 and then to just 2 trials (3%) in 2002 (Table II). There was a preference of authors of certain types of surgical trial (e.g. different methods of insemination/embryo transfer) to use crossover designs. Of the 21 crossover trials reviewed, 19 may have been adequately randomized (of which only five provide details of the randomization, two of which also had adequately concealed the allocation schedule). Only two crossover trials were classified as possibly randomized and reported pre-crossover pregnancy data.

**Patient flow**

As a minimum, trials that are identified as suitable for inclusion in Cochrane Reviews should have adequate randomization of patients to interventions. Often this is interpreted as merely stating that patients were randomly assigned to the treatments, without further detail, and there being no evidence of this being done in either a systematic or a haphazard way. Of the

---

**Table I. Number of trials available for review**

<table>
<thead>
<tr>
<th>Journal</th>
<th>Year</th>
<th>Specialty of main intervention</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1990</td>
<td>1996</td>
<td>2002</td>
</tr>
<tr>
<td>Fertility &amp; Sterility</td>
<td>21</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>Human Reproduction</td>
<td>12</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Other journals</td>
<td>12</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>60</td>
<td>59</td>
</tr>
</tbody>
</table>

*aOne acupuncture trial, published in 2002, was included in this category.

**Figure 1. Numbers of trials providing ‘adequate’, ‘unclear’ or ‘inadequate’ descriptions for the randomization, allocation and study design stages.**
164 subfertility trials reviewed, we found that only 140 (85%) satisfied these criteria. Of these, only 99 (60%) provided details on the number of women initially randomized to each group, as opposed to just the number of women analysed in each group, eight of which did not adequately report the number of women in each group achieving a pregnancy: three were crossover trials with first-stage data not extractable, three reported pregnancies only as percentages and the actual numbers were not extractable, one had contradictory numbers in the tables and text and the other stated only that there had been no increase in the pregnancy rate. Only 91 trials (55%) reported sufficient details on the number of women randomized and clinical pregnancies for each group; 57 also reported sufficient details on the number of ongoing pregnancies, of which 17 also reported sufficient details on the number of live births in each group. If strict quality criteria are applied, only 10 of the 164 (6%) trials assessed adequately reported on the randomization and allocation concealment methods, of which only three (all published in 2002, in HR and F&S) had sufficient details extractable to allow for an ITT analysis of the outcome ‘live birth’ (Figure 2).

**Pregnancy-related adverse events**

Of the 99 trials that may have used adequate randomization and reported sufficient details on the number of women randomized to each group, 58 had details extractable on the number of miscarriages for each group, 27 had details on the number of ectopic pregnancies and 36 had details on the number of multiple pregnancies in each group. A few reports mentioned the number of miscarriages and ectopic and multiple pregnancies in some way (e.g. as total number of events in the trial or for certain subgroups), but results for each randomized group were not extractable (Table III). Twenty trials (20%) had details extractable for all three adverse events, and three reports did not mention any of them. Miscarriages were fairly widely reported in the leading subfertility journals (F&S and HR), but reporting of ectopic and multiple pregnancies was less common.

**Discussion**

The effect of the CONSORT statement on the quality of RCTs has been assessed in the past: Moher et al. (2001a) concluded that the publication of CONSORT had had an impact on certain aspects of the quality of reporting of RCTs by 1998, when compared with 1994. Devereaux et al. (2002) found that the promotion of the CONSORT checklist to authors was associated with better reporting. Other studies have suggested that the awareness of the problems of poor reporting of medical research (Altman, 1994; Schulz et al., 1994, 1995) was already pushing up the quality of RCT reports before the publication of the CONSORT statement. Nevertheless, all studies concluded that there was still great scope for improvement in trial quality, particularly regarding the method of generating and concealing the treatment allocation sequence, which agrees with the results of the present review.

The proportion of subfertility trials found to have adequately reported random treatment allocation in 1990 and 1996 was 30%. These results are similar to those obtained in previous...
Figure 2. Flow chart of reporting quality.

Table III. Reporting of pregnancy related adverse events for trials suitable for inclusion in Cochrane Reviews

<table>
<thead>
<tr>
<th>Event</th>
<th>Year 1990</th>
<th>Year 1996</th>
<th>Year 2002</th>
<th>Specialty</th>
<th>Journal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laboratory</td>
<td>Medical</td>
<td>Surgical</td>
<td>Fertility &amp; Sterility</td>
<td>Human Reproduction</td>
<td>Other journals</td>
</tr>
<tr>
<td></td>
<td>Correctly or extractable</td>
<td>Incorrectly and not extractable</td>
<td>Not reported</td>
<td>Correctly or extractable</td>
<td>Incorrectly and not extractable</td>
<td>Not reported</td>
</tr>
<tr>
<td>Miscarriages [n (%)]</td>
<td>12 (52)</td>
<td>24 (60)</td>
<td>22 (61)</td>
<td>9 (60)</td>
<td>38 (60)</td>
<td>11 (52)</td>
</tr>
<tr>
<td>Correctly or extractable</td>
<td>0 (0)</td>
<td>3 (8)</td>
<td>3 (8)</td>
<td>2 (13)</td>
<td>2 (3)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Incorrectly and not extractable</td>
<td>11 (48)</td>
<td>13 (33)</td>
<td>11 (31)</td>
<td>4 (27)</td>
<td>23 (37)</td>
<td>8 (38)</td>
</tr>
<tr>
<td>Not reported</td>
<td>23 (100)</td>
<td>40 (100)</td>
<td>36 (100)</td>
<td>15 (100)</td>
<td>63 (100)</td>
<td>21 (100)</td>
</tr>
<tr>
<td>Ectopic pregnancies [n (%)]</td>
<td>6 (26)</td>
<td>11 (28)</td>
<td>10 (28)</td>
<td>0 (0)</td>
<td>20 (33)</td>
<td>7 (33)</td>
</tr>
<tr>
<td>Correctly or extractable</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Incorrectly and not extractable</td>
<td>17 (74)</td>
<td>28 (70)</td>
<td>26 (72)</td>
<td>15 (100)</td>
<td>42 (68)</td>
<td>13 (62)</td>
</tr>
<tr>
<td>Not reported</td>
<td>23 (100)</td>
<td>40 (100)</td>
<td>36 (100)</td>
<td>15 (100)</td>
<td>63 (100)</td>
<td>21 (100)</td>
</tr>
<tr>
<td>Multiple pregnancies [n (%)]</td>
<td>5 (22)</td>
<td>15 (38)</td>
<td>16 (44)</td>
<td>7 (47)</td>
<td>21 (33)</td>
<td>8 (38)</td>
</tr>
<tr>
<td>Correctly or extractable</td>
<td>2 (9)</td>
<td>4 (10)</td>
<td>6 (17)</td>
<td>3 (20)</td>
<td>6 (10)</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Incorrectly and not extractable</td>
<td>16 (70)</td>
<td>21 (53)</td>
<td>14 (39)</td>
<td>5 (33)</td>
<td>36 (57)</td>
<td>10 (48)</td>
</tr>
<tr>
<td>Not reported</td>
<td>23 (100)</td>
<td>40 (100)</td>
<td>36 (100)</td>
<td>15 (100)</td>
<td>63 (100)</td>
<td>21 (100)</td>
</tr>
</tbody>
</table>

*One acupuncture trial, published in 2002, was included in this category.
pre-CONSORT reviews of subfertility and obstetrics and gynaecology, where 29 and 32% of references were found to have adequately reported the randomization method (Vandenkerckhove et al., 1993; Schulz et al., 1994), but lower than the 49% (Altman and Doré, 1990) and 52% (Kjaergard et al., 1999) from reviews of trials published only in high-impact journals. In 2002, 56% of trials published in F&S and HR adequately reported a random method of treatment allocation compared with 64% of trials published in 2001 (Vail and Gardener, 2003). Reviews of the quality of trials published after CONSORT have found proportions of 23–60% in general medical fields (Devereaux et al., 2002; Hill et al., 2002), setting subfertility trials in the upper half of this range.

We have seen that 17% of pre-CONSORT and 37% of post-CONSORT trials mention some form of allocation concealment (even if not completely clear). Considering only the main subfertility journals (F&S and HR), we obtain 20 and 40% as pre- and post-CONSORT proportions, respectively. This is comparable to other results in the literature: 23–32%, pre-CONSORT (Schulz et al., 1994, 1995, 1996) and 33% after CONSORT (Vail and Gardener, 2003).

In the years included in this review, the proportions of trials reporting both adequate randomization and allocation concealment were 10% (6/59) after CONSORT and 4% (4/105) before CONSORT. If we include reports with unclear descriptions of allocation concealment, these proportions rise to 9 and 19% before and after CONSORT, respectively, rising again to 10 and 24%, respectively, when only the two main subfertility journals are considered. The overall proportion of the reviewed trials reporting adequate randomization and allocation concealment was 6% (10/164).

Of the aspects specific to subfertility, it is noteworthy that the number (and proportion) of crossover trials published in each year decreased. However, reporting of pregnancy-related adverse events, the timing of randomization and the number of women randomized to each treatment is still poor.

The unit-of-analysis errors were again identified as a major cause of poor quality in the subfertility literature. This has been noted as a common problem in other areas, such as ophthalmology (Newcombe and Duff, 1987) or dentistry (Blomquist, 1985), where each patient contributes multiple observations (two eyes and multiple teeth) and is covered in the BMJ’s Statistics Notes (Altman and Bland, 1997). Survival analysis, hierarchical models and other statistical methods have been developed to adequately analyse studies where each patient contributes multiple observations. In the area of subfertility, there is the added complication of informative censoring: reasons for not participating in subsequent cycles may depend on prognosis following a first unsuccessful cycle. It is therefore not clear whether the direct application of methods developed for other areas would yield valid analyses in subfertility, and this is currently the subject of ongoing statistical research.

In this review, only 10 trials provided evidence of having adequately allocated patients to interventions, of which only three had extractable details on the number of women randomized, ongoing pregnancies and the number of births in each group. Given that poor reporting is often associated with poor methodological quality (Schulz et al., 1995), much still needs to be done to improve the quality of RCTs in subfertility. Although some of the issues detected in this review relate to fundamental design flaws, others are exclusively due to poor reporting. This suggests that although many journals claim to have adopted the CONSORT statement, referees and editors need to be more aware of its requirements.

Our concern is the ability to include data from published trials in Cochrane Reviews of relevant interventions for subfertility. The quality criteria of Cochrane Reviews of RCTs were taken as the basis for our own definition of quality. We have found that a general improvement in reporting quality would mean that meta-analyses would be substantially more informative since fewer trials would be excluded because of the lack of information being provided. In the meantime, systematic reviewers will need to continue contacting authors for more detailed information.

Acknowledgements
The authors acknowledge the help of Michelle Proctor and Lisa McComb-Williams from the Cochrane MDSG base in Auckland, New Zealand, in accessing the trial database. This work has been funded by a Wellbeing of Women (RCOG) grant.

Conflicts of interest
Sofia Dias and Andy Vail are statistical editors for the Cochrane Menstrual Disorders and Subfertility Group.

References
Improving quality in subfertility RCTs


